Biophysical Journal, Volume *117*

Supplemental Information

On the Interpretation of Force-Induced Unfolding Studies of Membrane Proteins Using Fast Simulations

Zongan Wang, John M. Jumper, Karl F. Freed, and Tobin R. Sosnick

Supplemental Information for

On the interpretation of force-induced unfolding studies of membrane proteins using fast simulations

Zongan Wang[1,2], John M. Jumper[1,2,3,5], Karl F. Freed[1,2,*], Tobin R. Sosnick[2,3,4,*]
[1]Department of Chemistry, [2]James Franck Institute, [3]Department of Biochemistry and Molecular Biology, [4]Institute for Biophysical Dynamics, The University of Chicago, Chicago, IL, 60637.
[5]Current address: DeepMind, 5 New Street Square, London EC4A 3TW, UK.
[*]To whom correspondence may be addressed.
Email: trsosnic@uchicago.edu or freed@uchicago.edu

This PDF file includes:

- Supplemental information text
- Supplemental information methods
- Figures S1 to S14
- Tables S1 to S7
- Caption for movie S1
- References for SI reference citations

Supplemental Information Text

Other SMFS modes applied to GlpG.
We also performed standard force clamp simulations where the force is rapidly set and held at a constant value throughout the unfolding trajectory. The values generally are set at a force substantially less than the level where the first unfolding event occurs when operating under pulling mode with increasing force. We find that multi-step sequential unfolding from both N-to-C and C-to-N are more likely to be seen under lower force (e.g., 40 versus 60 pN), though the protein tends to unfold more cooperatively and more quickly at either force compared to the pulling with the stiff cantilever (**Figs. S12D and S13**).

Altering the pathway fluxes using mutation, temperature and spring constant.
We also performed unfolding simulations on destabilizing GlpG mutants to examine the effects on the unfolding pathways. Of the investigated residues having an H-bonding side chain in the N-domain, the E166A mutation is the most destabilizing (1). This residue is located near the bottom of the TM2 helix and forms two H-bonds to the backbone nitrogens of Val96 and Thr97 on the TM1 helix and two to the side chains of Thr97 on the TM1 and Ser171 on the TM3 helix (1). The G261V mutation on helix TM6 is at the center of the GxxxGxxxA motif that enables the close backbone-backbone association of the TM4 and TM6 helices. This mutation decreases the T_m by 28.1±0.08 °C and increases the probability of unfolding from the C-domain by 50% (10 to 15 events, of a total of 50) (**Fig. S12A and Table S3**). To our surprise, the disruption of the H-bond network at the bottom of the triad of the three TM helices in the N-domain barely changes the probability of initiating the unfolding from this end (40 versus 41 events of a total of 50) (**Fig. S12A and Table S3**).

Unfolding from the N- rather than the C-terminus is 4-fold more probable at 270 K. At 300 K, the ratio is reduced to 1.2. And weakening the spring constant by a factor of 5 (0.01 to 0.05 $k_BT/Å^2$) further reduces the branching ratio to 0.7 (**Fig. S9 and Table S3**). Even though there are differences, the fundamental heterogeneous pathway behavior remains (**Fig. S12A**).

Supplemental Information Methods

Calibration of virtual cantilever using thermal fluctuations.
To test whether our spring constant, κ, functions as intended, we compared the observed thermal fluctuations of the cantilever to those expected from the equipartition theorem, $\langle z^2 \rangle = k_B T/\kappa$ (2). We used the first 3, 10, 20, 50 residues of bR and ran simulations with the first residue attached to the virtual cantilever and the rest of the segment restrained as a rigid body. In this case, we can measure the thermal fluctuation of the tip of the cantilever via the fluctuation of the residue (**Fig. S1A**). The square root of the mean fluctuations has a linear relation with the reverse of the square root of the spring constant (**Fig. S1B**) (2).

Worm-like chain (WLC) model and the analytical solution to contour length
Unfolded proteins and nucleic acids behavior under force can be described with a worm-like chain (WLC) model for polymer elasticity (3-5). We present an analytic solution for the contour length Lc as a function of force and extension. In the WLC model (3), the force F and the extension x of the unfolded protein has the following relation:

$$F = \frac{k_B T}{L_p}\left[\frac{1}{4}\left(1-\frac{x}{L_c}\right)^{-2} + \frac{x}{L_c} - \frac{1}{4}\right] \qquad (1)$$

where k_B is the Boltzmann constant, T is the temperature, Lp = 0.4 nm is the persistent length of unfolded polypeptide, and Lc is the contour length (total length) of the unfolded polypeptide.

Let $\alpha = \frac{k_B T}{L_p}$, $\lambda = 1 - \frac{x}{L_c}$, $\omega = \frac{4F}{\alpha} - 3$ and substitute them into **Eq. 1**, we have

$$4\lambda^3 + \omega\lambda^2 - 1 = 0 \,(2)$$

According to Cardano's method (6), any cubic equation can be solved analytically: $ax^3 + bx^2 + cx + d = 0$ $(a \neq 0, a, b, c, d \in \mathbb{R})$.

Let $x = y - \frac{b}{3a} \Rightarrow y^3 + \left(-\frac{b^2}{3a^2} + \frac{c}{a}\right)y + \left(\frac{2b^3}{27a^3} - \frac{bc}{3a^2} + \frac{d}{a}\right) = 0$

Let $\begin{cases} P = -\frac{b^2}{3a^2} + \frac{c}{a} \\ Q = \frac{2b^3}{27a^3} - \frac{bc}{3a^2} + \frac{d}{a} \end{cases} \Rightarrow y^3 + Py + Q = 0$

Let $\Delta = \left(\frac{P}{3}\right)^3 + \left(\frac{Q}{2}\right)^2$ and $\begin{cases} S = \left(-\frac{Q}{2} + \sqrt{\Delta}\right)^{1/3} \\ T = \left(-\frac{Q}{2} - \sqrt{\Delta}\right)^{1/3} \end{cases}$, we have three roots: $\begin{cases} y_1 = S + T \\ y_2 = \beta S + \beta^2 T \\ y_3 = \beta^2 S + \beta T \end{cases}$, where

$\beta = \frac{-1 + i\sqrt{3}}{2}$ and $\beta^2 = \frac{-1 - i\sqrt{3}}{2}$ are the two complex cubic roots of -1.

Here, Δ is the discriminant of the cubic equation.

If $\Delta > 0$, there is only one real root, y_1 and two complex roots y_2 and y_3.
If $\Delta = 0$, if P = Q = 0 all three roots are equal to 0, otherwise there are three real roots and two of them are equal.
If $\Delta < 0$, there are three unequal real roots with the following relation:

$$\begin{cases} x_1 + x_2 + x_3 = -\dfrac{b}{a} \\ 1/x_1 + 1/x_2 + 1/x_3 = -\dfrac{c}{d}, \text{ where } x_i = y_i - \dfrac{b}{3a}, \text{ i = 1, 2, 3.} \\ x_1 \bullet x_2 \bullet x_3 = -\dfrac{d}{a} \end{cases}$$

Now, back to **Eq. 2**, let $\lambda = y - \dfrac{\omega}{12}$ (3), we have

$$y^3 + \frac{-\omega^2}{48} y + \left(\frac{\omega^3}{27*32} - \frac{1}{4} \right) = 0 \text{ (4), and}$$

$$\Delta = \left(\frac{-\omega^2}{48*3} \right)^3 + \left(\frac{\omega^3}{27*64} - \frac{1}{8} \right)^2 = \frac{1}{64} \left(1 - \frac{\omega^3}{27*4} \right) \text{ (5).}$$

In a standard SMFS experiment of unfolding bR, the force F is between 0 and 500 pN. At T = 298 K, k_BT = 4.114 pN•nm. Only when F < 20 pN is $\Delta > 0$; otherwise $\Delta < 0$. Therefore, **Eq. 2** has only one real root mostly (F ≥ 20 pN, $\Delta > 0$), which is the solution to our problem.

When $\Delta < 0$, we have

$$\begin{cases} \lambda_1 + \lambda_2 + \lambda_3 = -\dfrac{\omega}{4} < 0 \\ 1/\lambda_1 + 1/\lambda_2 + 1/\lambda_3 = 0 \quad (6). \\ \lambda_1 \bullet \lambda_2 \bullet \lambda_3 = \dfrac{1}{4} > 0 \end{cases}$$

Assuming $\lambda_1 \le \lambda_2 \le \lambda_3$, we have $\lambda_1 < \lambda_2 < 0 < \lambda_3$ and λ_3 is the root we want. In summary, Lc can be solved analytically given force and extension.

Unfolding pathway analysis of bR.
For every frame in trajectory, the Lc of the already unfolded segment can be determined through either FEC or TSS. Assuming intact secondary structure remains unchanged within the bilayer, Lc is uniquely determined given a force and an extension (labeled as Lc,FEC), from which we can infer how many residues have unfolded. Force was measured and recorded into the H5 file during the simulation, while extension was calculated as the distance that the Cα atom of the C-terminus has moved. On the other hand, if the number of unfolded residues is known first, Lc can be determined by mapping the number of unfolded residues to pre-determined Lc value (labeled as Lc,TSS) (**Table S6**). Secondary structures were computed by the `compute_dssp` function

in *MDTraj* (9), which follows the DSSP definition (10). Then Lc,TSS is obtained after identifying the most C-terminal residue which remained folded.

Trajectories plotted according to the last (C-terminal) folded residue were smoothed by a Savitzky-Golay filter (11) in *Scipy* (12), in which the `window_length` was set to 11, `polyorder` 3, `mode` 'nearest'. Then, the population distribution was histogrammed and fitted with multiple Gaussian functions to identify the number position of the simulated intermediates. Amplitudes and positions were fit assuming a width (standard deviation) of one amino acid, i.e. the positional uncertainty is assumed to be ±1 amino acid. Three major unfolding regions, denoted ED, CB, and A, were fit separately. In general, the more intermediates, the smaller the fitting error. To prevent over-fitting, we initially assume that intermediates are evenly distributed within each major unfolding region and obtained the fitting error as a function of the number of intermediates. Later, by adjusting the number of intermediates and their initial positions manually, we acquired fewer intermediates with a relatively low fitting error (**Fig. S6**).

Calculation of contour length (Lc).
We measured Lc from simulations of truncated bR molecules. The truncation points were chosen to match the experimental intermediates (13). For example, we simulated "A160", a truncated version having residues 161-232 unfolded to match the intermediate where residues 1-160 are folded while 161-232 are unfolded. For each of the truncated bR species, we fit its FEC with a WLC model (**Eq. 1**) using a fixed persistence length (Lp) of 0.4 nm, estimated by experiment (14) (**Fig. S2A, B**), to determine Lc (**Fig. S2C and Tables S4, S6)**. From this plot of Lc values, we obtained an average slope of 0.390 nm•residue^{-1} in agreement with the experimental Lc estimate of 0.40±0.02 nm•residue^{-1} (15) (the average distance between consecutive Cα's is 0.38 nm for actual proteins). Our Lc value is ~7% larger than 0.364 nm•residue^{-1}, a value recently obtained by a high precision measurement (13). Remarkably, for the truncated bR molecules, our Lc values exhibit the same minor deviations from linearity as those observed experimentally. The reproduction of these small deviations implies that they are real. The only reasonable source of the variability is a sequence dependent for Lc, consistent with experimental (15) and simulation (16) findings. Beyond providing support for the accuracy of our simulations, the residue dependence should be useful in identifying the sequence of the segment that is unfolded for a given Lc value.

Force clamp simulations of ubiquitin.
We unfolded ubiquitin (1ubq.pdb) to its fully extended state under high force (800 pN) and ran force clamp simulations with a constant force applied to both ends of the protein (procedure described in **Table S5**). We replicated all-atom MD results (16) in a few cpu-hours. Without force, the highly stretched polypeptide contracts considerably but remains extended under force as low as 10 pN (**Fig. S3A**). The distributions of (ϕ, ψ) angles and end-to-end distances at different forces were similar to those of the all-atom MD study (16) (**Fig. S3B, C**). Also, we obtained good fitting of the average end-to-end distances and the applied forces according to the WLC model (**Fig. S3E**). The Lc was determined by minimizing the least-squares fitting error (**Fig. S3D**).

Clustering analysis of GlpG unfolding trajectories.
We chose the Gaussian mixture algorithm implemented in *scikit-learn* (17) to perform the clustering analysis performed on the structures in each trajectory (**Figs. S10B and 11B**). A Gaussian mixture model is a probabilistic model that assumes all the data points are generated from a mixture of a finite number of Gaussian distributions, which is good for density estimation. The number of clusters was estimated and supplied to the program. The corresponding density for each training point was measured and the point with the maximal density was chosen as the center to represent the cluster. The cluster centers were considered as the intermediate states along the unfolding pathway.

Estimating *Upside* temperature and time scale
We use Verlet integration with a time unit of 0.009 *Upside* time steps and random number generator to implement the Langevin dynamics with a thermalization time scale of 0.135 time units. The time scale of thermalization (related to Langevin friction) is chosen to maximize the effective diffusion rate of chains while effectively controlling simulation temperature. As Langevin dynamics with any friction coefficient produces the same Boltzmann ensemble, we chose to maximize equilibration of our system rather than attempt to match a solvent viscosity.

The precise time and temperature scale of the *Upside* model is unclear because of the coarse-graining process. As compared to all-atom, explicit solvent simulations on a decapeptide using Charmm36, *Upside* is about 10,000-fold faster in part due to the lack of solvent and the smoothing of side chain interactions. This smoothing is likely to have a disproportionate effect for condensed structures as compared to extended structures. From the transition rates between folded and unfolded states (well-to-well barrier crossing process), we estimate the time unit for barrier crossing events to be ~1 psec. From transition rate between Ramachandran basins in the extended state (chain motions within a thermodynamic well), we estimate the time unit to be ~0.1 psec. As noted above, each time unit is 0.009 time steps. Irrespective of the issues with defining an absolute time scale, the equilibrium population distribution that determines the free energy is expected to be approximately correct, as well as the order of dynamical folding events.

Simulation configurations of gradual pulling simulations.
In addition to the standard configuration for running *Upside* folding simulation of small soluble proteins (18, 19), we implement our membrane burial potential, which dynamically accounts for the degree of side chain exposure to lipids (20) and a new pulling function in the *Upside* simulations.

(1) Prepare the initial protein structure in a pickle file format the input.
```
python PDB_to_initial_structure.py \
pdbname.pdb pdbname                \
--allow-unexpected-chain-breaks    \
--record-chain-breaks              \
--disable-recentering
```

(2) Prepare the H5 file for the simulation, which includes all simulation parameters.
```
python upside_config.py                                    \
  --output             pdbname.h5                          \
  --fasta              pdbname.fasta                       \
```

```
  --initial-structure  pdbname.initial.pkl                \
  --hbond-energy       $(cat UPSIDE_param_dir/ff_1/hbond) \
  --dynamic-rotamer-1body                                 \
  --rotamer-placement   UPSIDE_param_dir/ff_1/sidechain.h5\
  --rotamer-interaction UPSIDE_param_dir ff_1/sidechain.h5\
  --environment UPSIDE_param_dir/ff_1/environment.h5      \
  --rama-sheet-mixing-energy $(cat UPSIDE_param_dir/ff_1/sheet)\
  --rama-library  UPSIDE_param_dir/common/rama.dat        \
  --reference-state-rama
UPSIDE_param_dir/common/rama_reference.pkl                \
  --membrane-thickness   membrane_thickness               \
  --membrane-potential   membrane_potential_fpath         \
  --ask-before-using-AFM AFM_fpath                        \
  --AFM-time-initial     0
```

AFM_fpath is the path to the file that defines the residue to which force will be applied, tip position, spring constant and pulling velocity. Force can be applied to one or more residues. In *Upside*, the unit of the energy is k_BT: 1 k_BT ≈ 4.114 pN•nm at T = 1.0 (≈ 298 K). The unit of the spring constant is k_BT/Å^2: 1 k_BT/Å^2 ≈ 41.14 pN/Å = 411.4 pN/nm. 1 *Upside* time step ≈ 0.1 ns, so the pulling velocity 0.001 Å/*Upside* time step ≈ 10^6 nm/s, the same as the extraction velocity in the CG-MD simulations (21).

(3) Run *Upside*.

```
upside pdbname.h5              \
  --seed           random_seed \
  --temperature    temperature \
  --frame-interval frame_intvl \
  --duration       duration    \
  --disable-recentering
```

Simulation configurations of force clamp simulations.

The only difference with the configuration above is in the preparation of the H5 file. A tension file is supplied to *Upside* instead of an AFM file, which defines the pulling residue and pulling force. One or more residues can be pulled.

```
python upside_config.py                                   \
  --output               pdbname.h5                       \
  --fasta                pdbname.fasta                    \
  --initial-structure    pdbname.initial.pkl              \
  --hbond-energy    $(cat UPSIDE_param_dir/ff_1/hbond)    \
  --dynamic-rotamer-1body                                 \
  --rotamer-placement UPSIDE_param_dir /ff_1/sidechain.h5 \
  --rotamer-interaction UPSIDE_param_dir ff_1/sidechain.h5\
  --environment       UPSIDE_param_dir/ff_1/environment.h5\
  --rama-sheet-mixing-energy $(cat UPSIDE_param_dir/ff_1/sheet)\
  --rama-library   UPSIDE_param_dir/common/rama.dat       \
  --reference-state-rama
UPSIDE_param_dir/common/rama_reference.pkl                \
```

```
  --membrane-thickness   membrane_thickness              \
  --membrane-potential   membrane_potential_fpath        \
  --tension              tension_fpath
```

Fasta sequences of the proteins used in the study.

1. bR (1qhj.pdb)

```
QAQITGRPEWIWLALGTALMGLGTLYFLVKGMGVSDPDAKKFYAITTLVPAIAFTMYLSMLLGY
GLTMVPFGGEQNPIYWARYADWLFTTPLLLLDLALLVDADQGTILALVGADGIMIGTGLVGALT
KVYSYRFVWWAISTAAMLYILYVLFFGFTSKAESMRPEVASTFKVLRNVTVVLWSAYPVVWLIG
SEGAGIVPLNIETLLFMVLDVSAKVGFGLILLRSRAIFGEAEAPEPSAGDGAAATS
```

2. GlpG (2xov.pdb)

```
ERAGPVTWVMMIACVVVFIAMQILGDQEVMLWLAWPFDPTLKFEFWRYFTHALMHFSLMHILFN
LLWWWYLGGAVEKRLGSGKLIVITLISALLSGYVQQKFSGPWFGGLSGVVYALMGYVWLRGERD
PQSGIYLQRGLIIFALIWIVAGWFDLFGMSMANGAHIAGLAVGLAMAFVDSLN
```

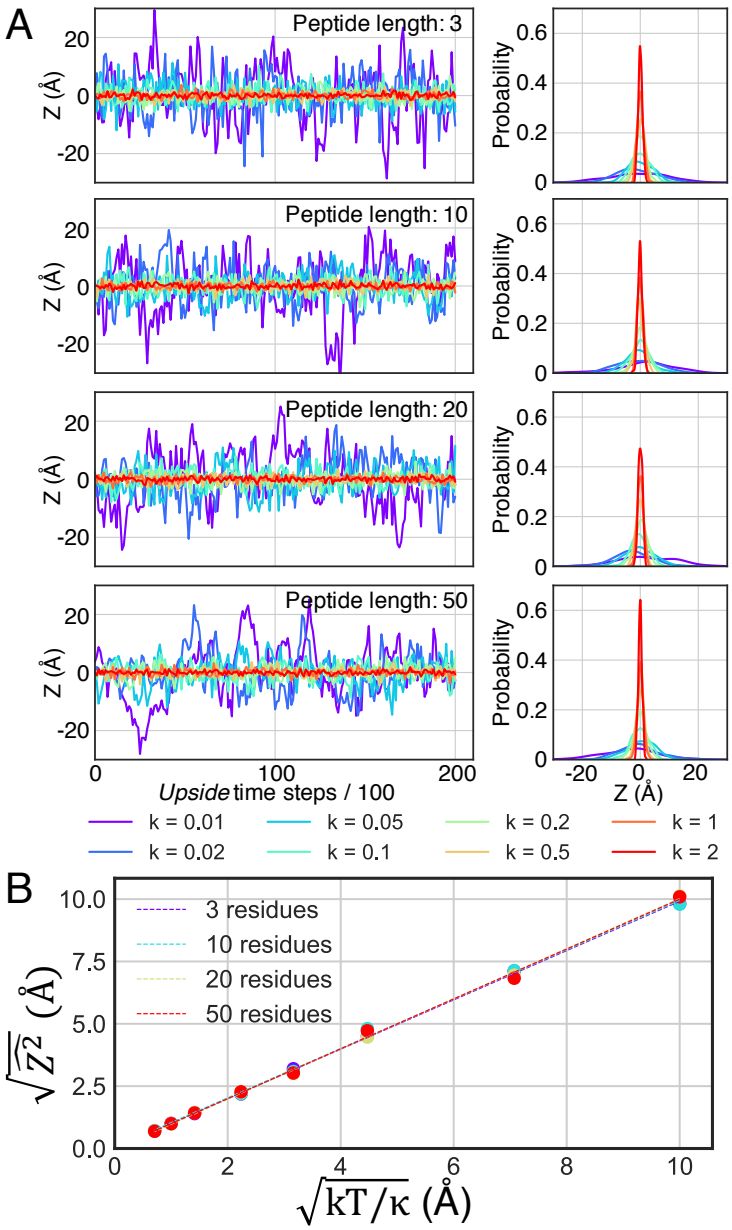

Fig. S1. Calibrating our virtual cantilever: Stiffness, thermal fluctuations, and the equipartition theorem. A. Thermal fluctuations and their distributions. A 3-50 residue segment of bR is attached to the tip of the cantilever and the fluctuations of the residue attached at the end to the cantilever are measured (i.e., same location, but with varying mass). **B.** In agreement with the equipartition theorem, $\langle z^2 \rangle = k_B T/\kappa$. The spring constant ($\kappa$) is in *Upside* unit: $k_B T/Å^2$.

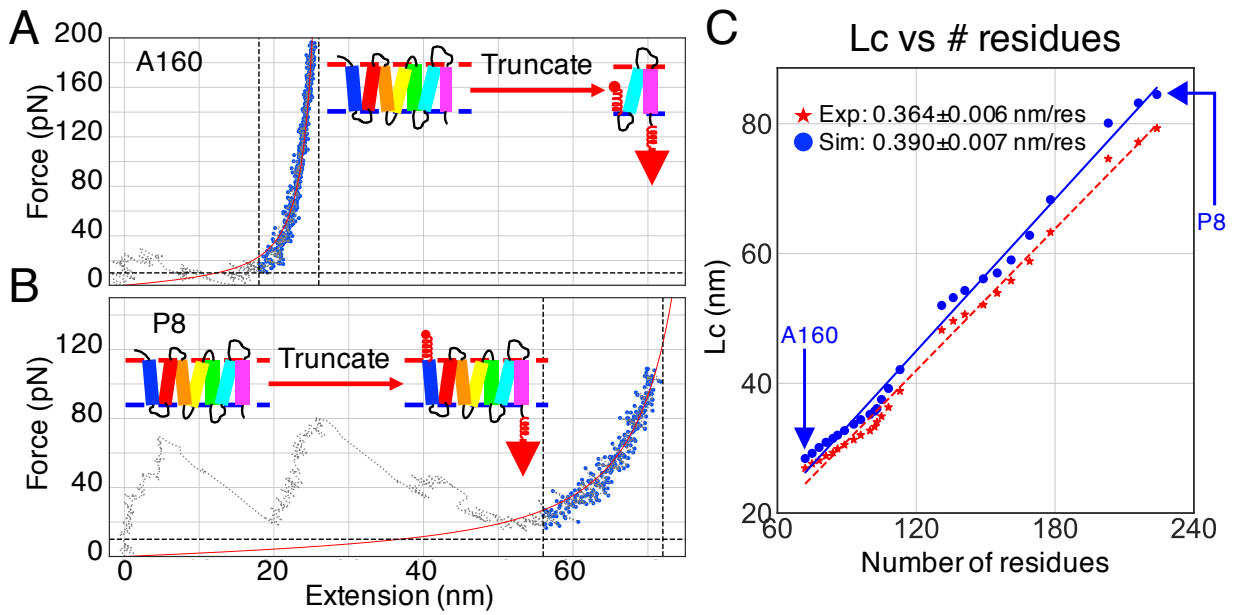

Fig. S2. Calibration of contour length (Lc) per amino acid. A, B. WLC fitting of the end-to-end distance (extension) and the force using a fixed Lp of 0.4 nm of truncated bR species A160 and P8, respectively. WLC fitting (curve in red) was performed on the data points in blue between the vertical black dashed lines (≥ 10 pN). The elasticity of the unfolded segment is well described by the WLC model. Note, during the initial (low force) portion, helices are being pulled out so the WLC is not applicable. Specifically, in **panel A**, the grey portion of the force-extension curve depicts the unfolding and pulling out of helix pair GF (similarly as shown in **Figs. 2B and S5**, at extension between 0 ~ 20 nm). After both helices G and F have been pulled out, the spring attached at the N-terminus (residue A160 in this case) starts getting extended, and thereafter we can see the WLC behavior (blue dots). **C.** Lc of unfolded segment as a function of number of residues from simulations compared to experiment (13). In the experiment, the number of unfolded amino acids is calculated based on $n_{aa} = (\Delta L_0 + \Delta d)/L_0^{aa}$, where Δd is the vertical distance of the folded structure along the pulling axis in native bR(22) and $L_0^{aa} = 0.366$ nm is the Lc per amino acid based the distance between the 1st intermediate in the helix pair ED (A160) and the 1st in the helix A (V29). The same deviations from linearity are observed in the simulations and experiment. This similarity implies that there is a similar sequence dependence that both highlights the accuracy of the simulations, and that the residue dependence could be useful in identifying the sequence of the segment that is unfolded for a given Lc value.

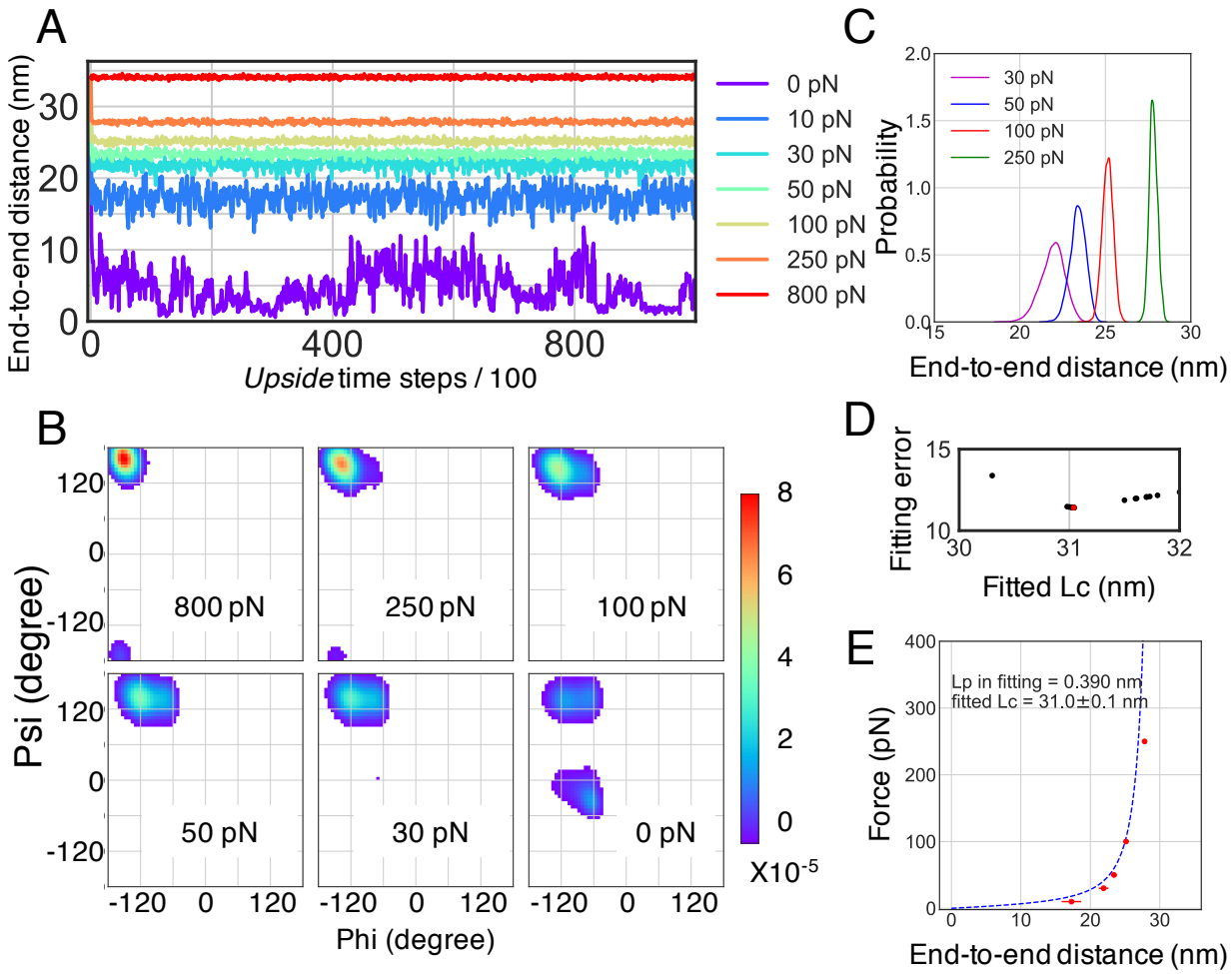

Fig. S3. Reproduction of the all-atom MD of unfolded ubiquitin. A. End-to-end distances of the protein under force. **B.** Distribution of ϕ and ψ dihedral angles of all 76 residues of the protein over time (in comparison to Fig. 2A of ref. (16)). Noticeably, there are populations of (ϕ, ψ) in the α-helix and left-handed helix regions under lower forces (100 pN, 50 pN, and 30 pN) in ref. (16). This is because of a specific version of force filed was used in the that study, in which the dihedral potentials are modified (23). In other words, our (ϕ, ψ) distributions would agree with ref. (16) if we had tuned our dihedral potentials. Nonetheless, (not) tuning the dihedral potentials does not affect the non-bonded interactions between side chains, as pointed out by ref. (16) (in the section "modified dihedral potentials" in SI of the paper), hence affect the end-to-end distances and the contour length. **C.** Probability density distribution of end-to-end distances (in comparison to Fig. S2 of ref. (16)). **D.** Fitting error versus fitted Lc. The minimum fitting error was obtained at Lc = 31.0 nm (red dot). **E.** WLC fitting of the average end-to-end distance and the applied force (in comparison to Fig. 1C of ref. (16)). The error bars show the standard deviation of the end-to-end distances. When fitting the data to obtain the Lc value, the value of Lp was fixed at 0.39 nm. Our fitted value for Lc is 31.0 nm, ~ 9% larger than 28.4 nm in ref. (16). The deviation of the force-extension to the WLC fitting is bigger at lower force (< 50pN) due to the interactions of non-neighboring residues when the protein collapses under low force, in which regime the chain behavior cannot be described by ideal chain models.

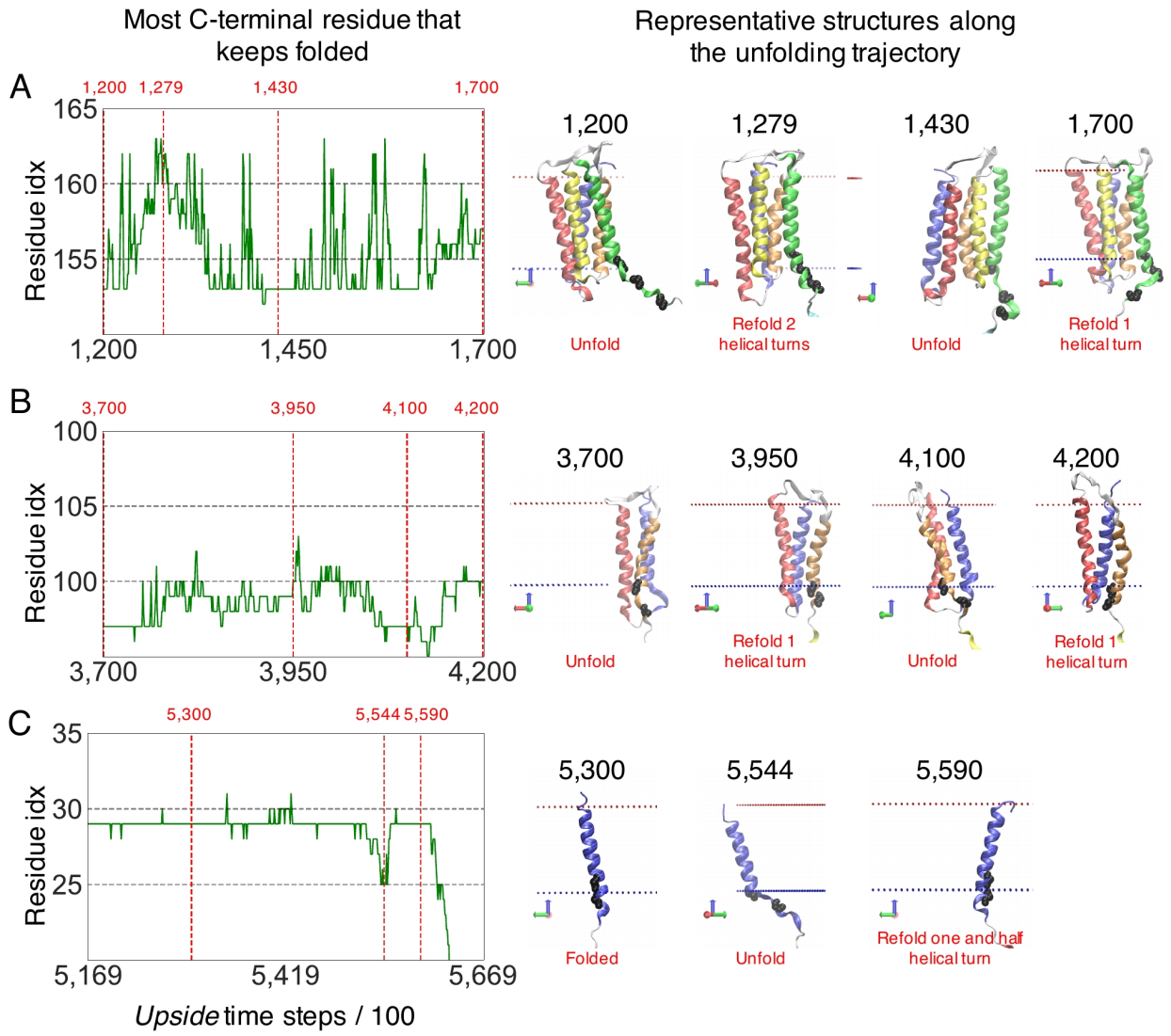

Fig. S4. Near equilibrium back-and-forth transitions between intermediates in helix pair ED (panel A), helix pair CB (panel B), and helix A (panel C). The inset in **panel C** of **Fig. 2** is enlarged and shown in the left column. In addition, one segment of TSS in the unfolding regions ED (from frame number 1200 to 1700) and one segment in A (from frame number 5169 to 5669) are shown. Representative structures are selected at the red vertical dashed lines and shown in the right column. The frame indices of those structures are above each plot. The TM helices are colored in the same set of codes as in **Fig. 1A**. Residues 161, 157 and 153 in structures in **panel A**, residues 100 and 97 in **panel B**, and residues 29 and 25 in **panel C** are plotted in the VDW presentation in black. The rapid back-and-forth transitions, which are considered as a hallmark of near-equilibrium measurement (13), can be seen in all three major unfolding regions in our simulations.

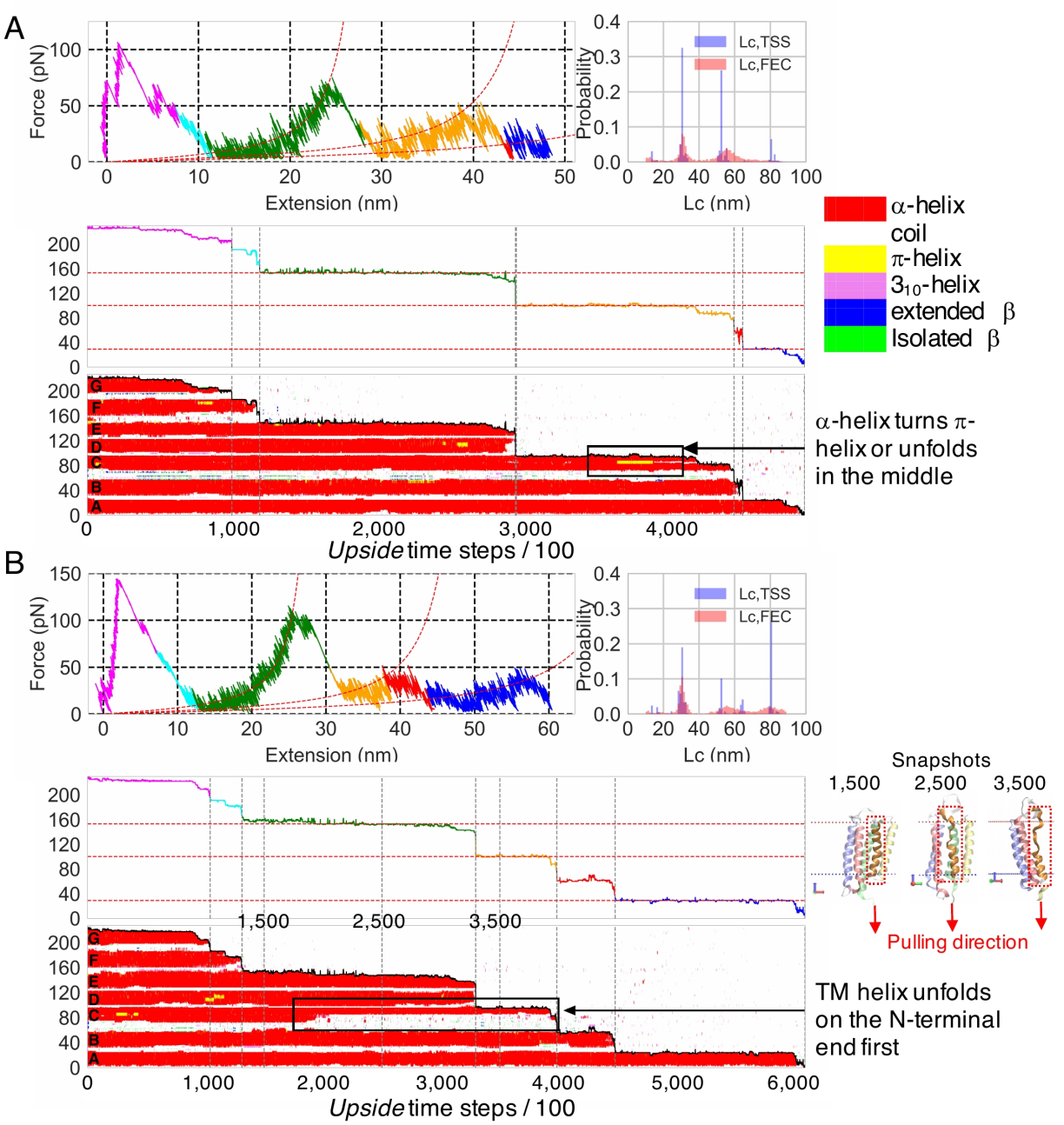

Fig. S5. Unfolding trajectories of bR. The trajectories largely support the common assumption that secondary structures remain intact within the membrane bilayer during the unfolding process. However, exceptions can be seen in these two TSS: **A.** Part of a TM helix may turn into π-helix or unfold in the middle (e.g., Helix C, black box in the TSS plot). **B.** A TM helix can unfold from the N-terminal end rather than the C-terminal end (e.g., Helix C, black box in the TSS plot). Snapshots taken at the 1500th, 2500th, are 3500th frames are shown. Helix C in orange (in red dashed box) unfolds at its N-terminal end first.

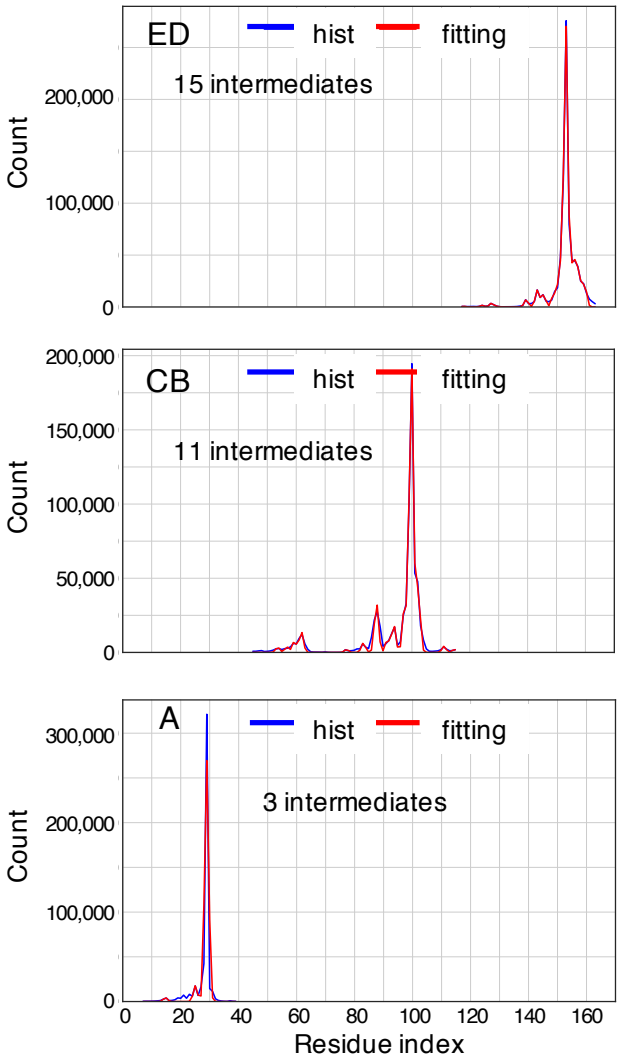

Fig. S6. Identifying intermediates by fitting with multiple Gaussian functions. For the 48 trajectories shown in **Fig. 3A**, the time spent at each position is histogrammed (blue lines) and fit using multiple Gaussians (red) with standard deviation (σ) of one residue to identify the number and position of intermediates. The number and initial position of the Gaussians was manually adjusted to minimize the fitting error; additional Gaussians were added until the error plateaued. The upper, middle and lower panels refer to unfolding occurring within the ED, CB or A helices, respectively. The index refers to the last residue that remains folded, as identified in the TSS, and is listed in **Table S2**.

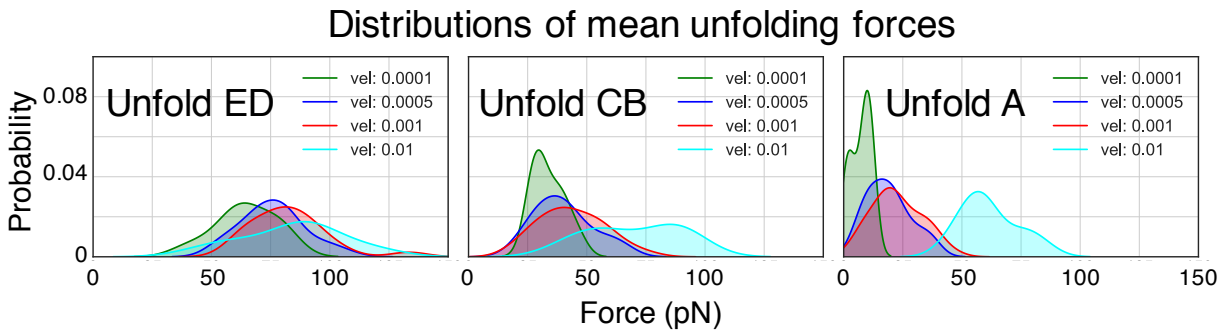

Fig. S7. Distributions of mean unfolding forces at different pulling velocities. The mean unfolding forces of the major intermediates in helix pairs ED, CB, and helix A (**Table S1**) are computed, respectively, from the trajectories simulated at each pulling velocity. Only the trajectories in which we observe all 7 helices are unfolded in the order ED, CB, and A, are included in the analysis. The velocities are in unit Å/*Upside* time step.

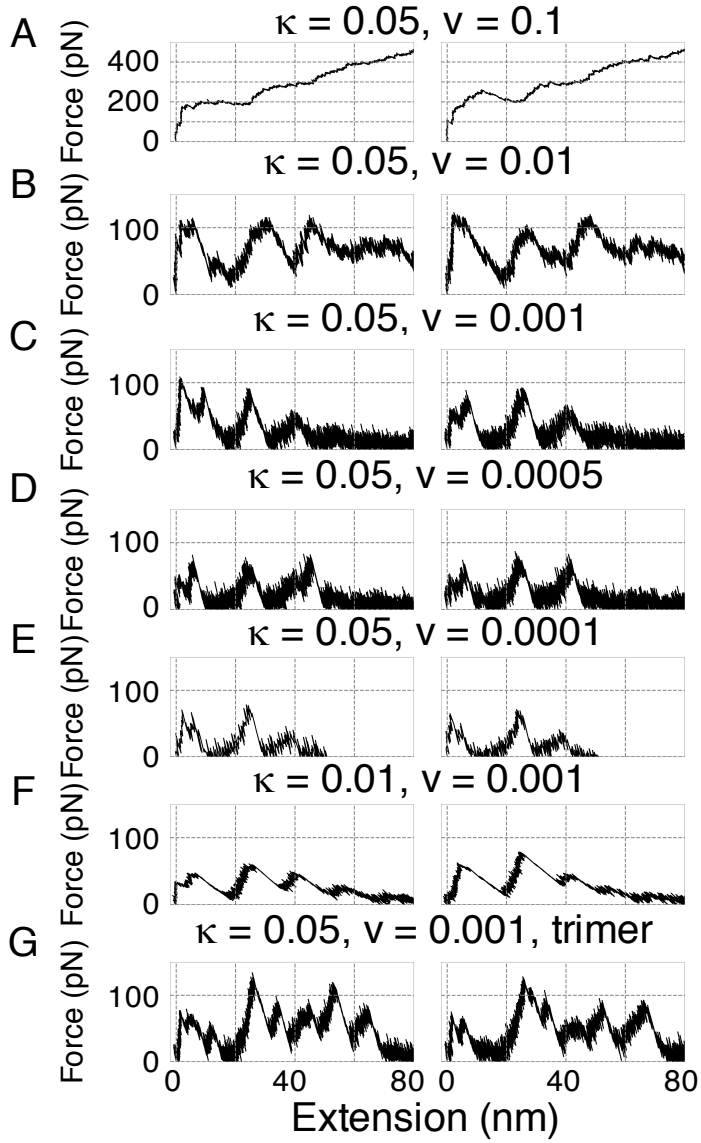

Fig. S8. Force-extension curves (FECs) of unfolding trajectories of monomeric (A to F) and trimeric bR (G) under different conditions. Two trajectories are shown for each set of simulation conditions. Details of the simulation setup are listed in **Table S5**. The unit of the spring constant is $k_BT/Å^2$: 1 $k_BT/Å^2$ ≈ 41.14 pN/Å = 411.4 pN/nm at T ≈ 300K. The pulling velocity 0.001 Å/*Upside* time step ≈ 10^6 nm/s. WLC behavior is not observed at the highest pulling velocity (**panel A**). At higher pulling velocity (**panels A, B**), force does not restore to zero after the entire protein is pulled out of the membrane bilayer due to the friction of the solvent. As the pulling velocity decreases (from 0.01 to 0.0001, **panels B to E**), the saw-tooth pattern in general has the same depth (the ruptures have similar slopes). When the spring constant decreases (from 0.05 to 0.01), the saw-tooth patterns become shallower, i.e. the slope of the rupture is decreased (**panels C, F**). More intermediates can be observed in the FECs of trimer (**panel G**, **Tables S2**).

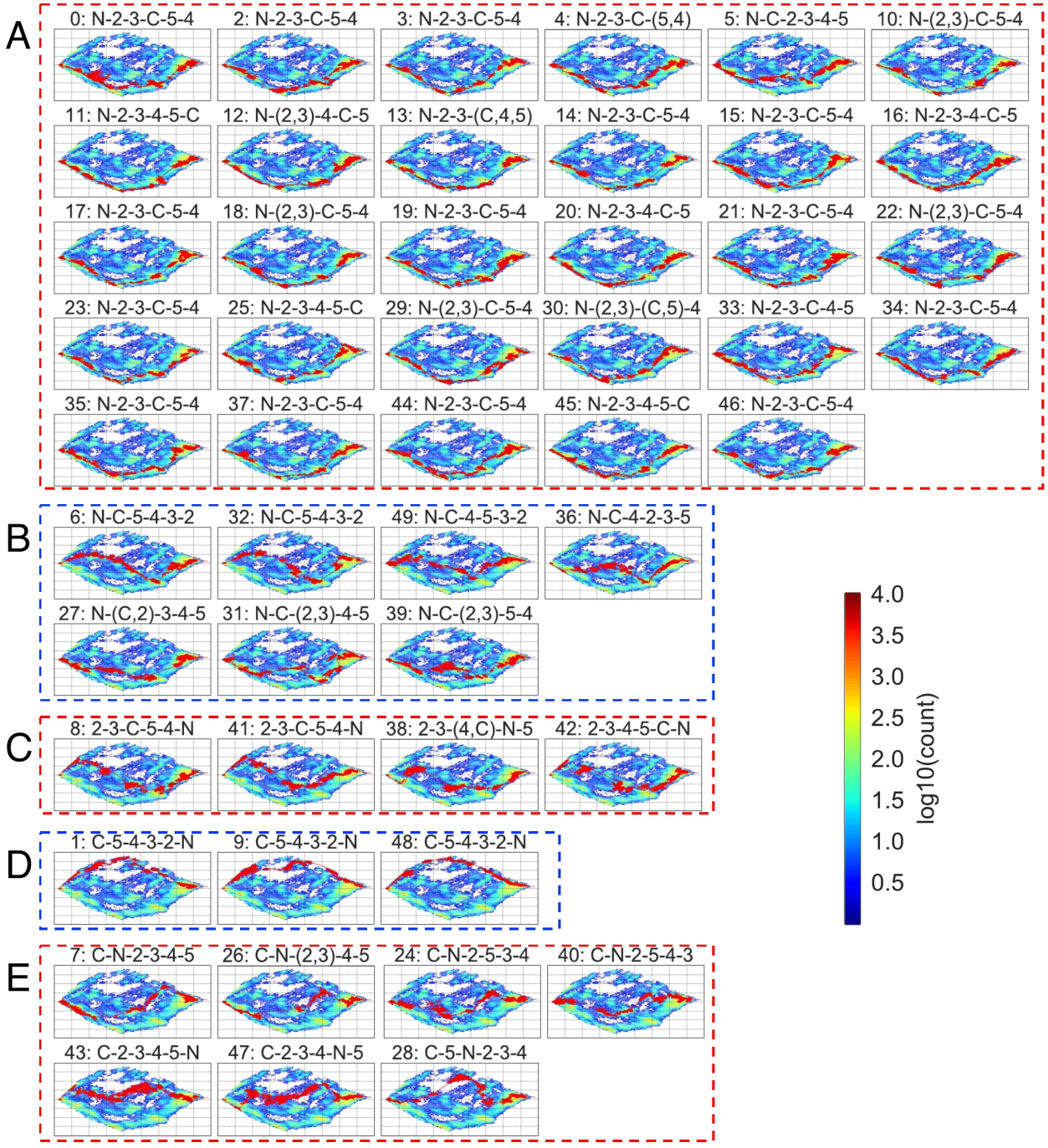

Fig. S9. Unfolding trajectories of GlpG obtained with a stiff cantilever (T = 0.9 ≈ 270 K, κ = 0.05 k_BT/Å^2, v = 0.001 Å/*Upside* time step; pulling the C-terminus and fixing the N-terminus with an equal strength spring). The heat map is obtained from 50 trajectories. The red curve of each subplot is the unfolding pathway from the native state to the fully extended (FE) state for a given trajectory. The trajectories are categorized based on their unfolding pathways. The title of each subplot indicates the index of the trajectory and its unfolding pathway. For example, "4: N-2-3-C-(5,4)" denotes the unfolding pathway of the fourth trajectory where the unfolding order is N→2→C→(5, 4), in which TM5 and TM4 unfold nearly simultaneously and therefore are put in parentheses.

A. Unfolding starts from TM1 (N) and proceeds to the C-domain when all the TM helices in the N-terminal domain unfold. The pathways traverse the lower edge of the PCA plots.

B. Unfolding starts from TM1, followed by the unfolding of TM6, leading to zigzag pathways across the PCA plots.

C. Unfolding starts from the middle of GlpG.

D. Unfolding starts from TM6 and proceeds to the N-domain when all the TM helices in the C-domain unfold. In contrast to **panel A**, the pathways flank the upper edge of the PCA plots.

E. Similar to **panel B**, with unfolding starting from TM6, followed by the unfolding of a TM helix in the N-domain, which results in zigzag pattern through the middle on the PCA plot.

The ratio of unfolding from the N-domain first to unfolding from the C-domain first is 40:10 (number of trajectories in panels **A**+**B**+**C** : number of trajectories in panels **D**+**E**).

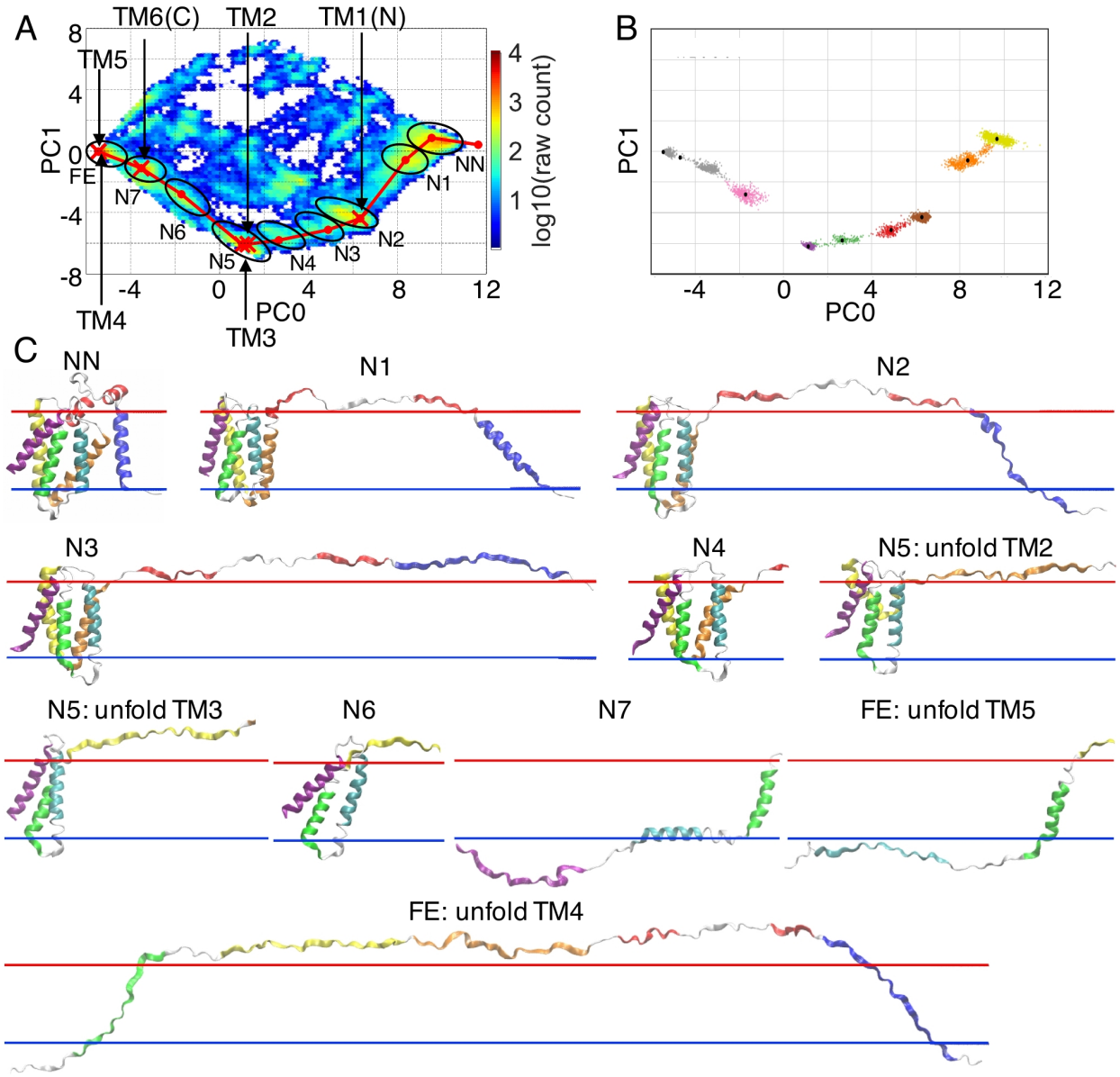

Fig. S10. Examples of an N→C unfolding pathway of GlpG. A. Unfolding pathway connected by representative structures on the PCA plot. The representative intermediates are chosen such that they are either the cluster center or the structure when a TM helix unfolds. These structures are considered as the intermediates. We use NN, N1, …, N7, and FE to denote the clusters as well as the intermediates. NN is short for near-native, and FE fully-extended. **B.** Clustering analysis of the trajectory. Nine clusters are identified. **C.** Snapshots of the representative structures. For the illustrative reasons, unfolded segments sometimes are not shown in the snapshots when there is no significant conformational change.

In the NN state, helices rearrange. In N1, the two interfacial helices H1, H2 unfold and separate. In N2, TM1 (N) unfolds. In N3, TM1 flips to the other side of bilayer. In N4, TM4 partially unfolds in its C-terminal. In N5, TM2 and TM3 unfold. In N6, the helices in the C-domain rearrange. In N7, TM6 (C) unfolds, and the C- terminus of TM4 refolds. In FE, TM5 and TM4 unfold, unfolded TM1 may re-enter the bilayer.

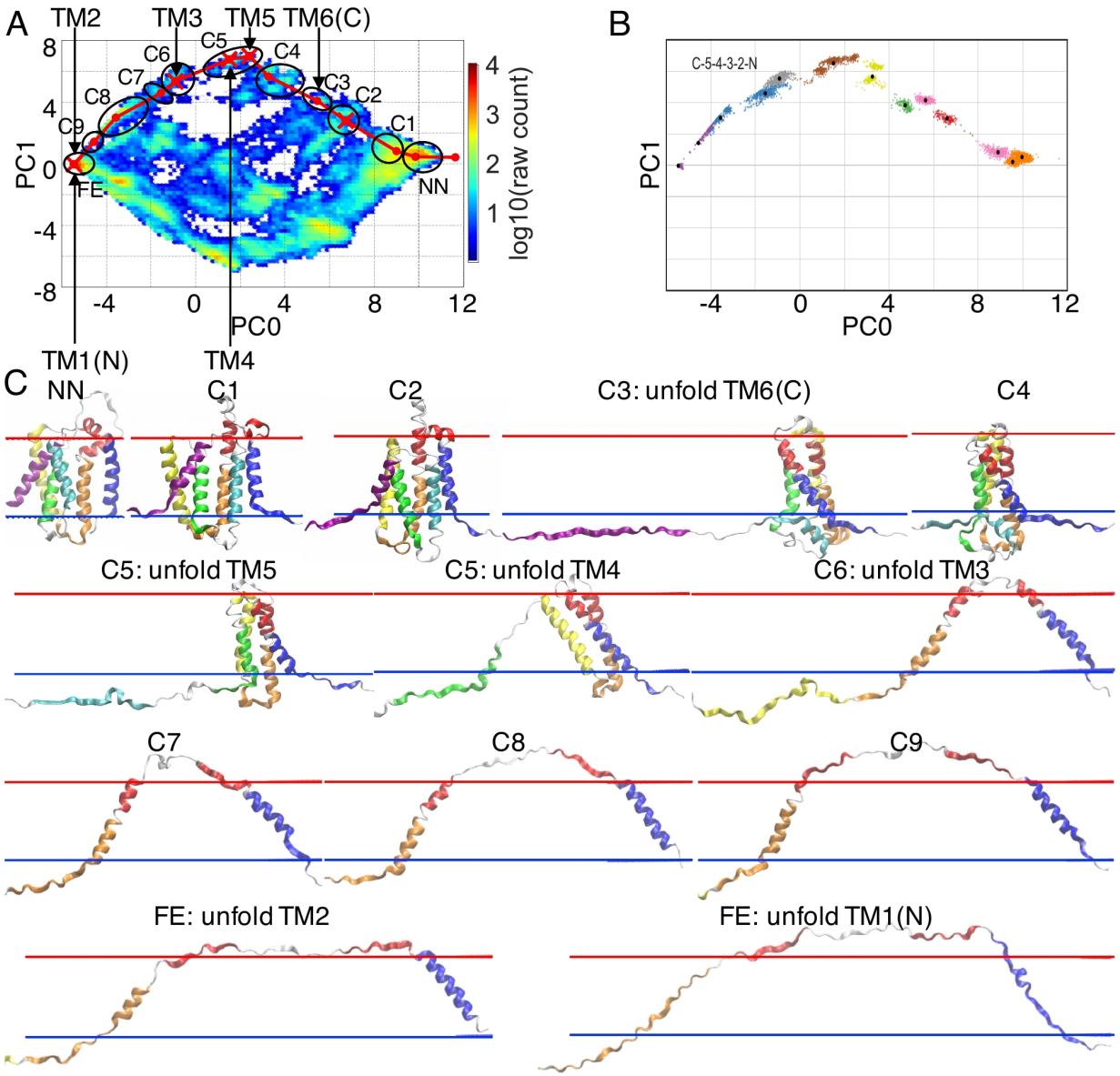

Fig. S11. Example of a C→N unfolding pathway of GlpG. A. Unfolding pathway connected by representative structures on the PCA plot**.** We use NN, C1, …, C9, and FE to denote the clusters as well as the intermediates. **B.** Clustering analysis. Thirteen clusters are identified. **c.** Snapshots of the representative structures.

In the NN state, helices rearrange. In C1, the interfacial helix H2 aligns with TM2 and pushes part of TM2 out of the bilayer, TM2 bends, TM1 (N) and TM6 (C) partially unfold, and H1 unfolds. In C2, TM6 unfolds one more helical turn, TM2 partially unfolds in its C-term and H1 refolds. In C3, TM6 unfolds, TM5 comes out of the bilayer, H1 aligns with TM1, and TM1 tilts in order to accommodate the elongation in its C-term due to the alignment of H1. In C4, two more helical turns of TM1 unfold, and TM4 and TM5 partially unfold. In C5, TM5 and TM4 unfold. In C6, TM3 unfolds, TM1 and TM2 come apart. In C7, H1 unfolds. In C8, TM1 and TM2 come further apart. In C9, H2 unfolds. In FE, TM2 and TM1 unfold.

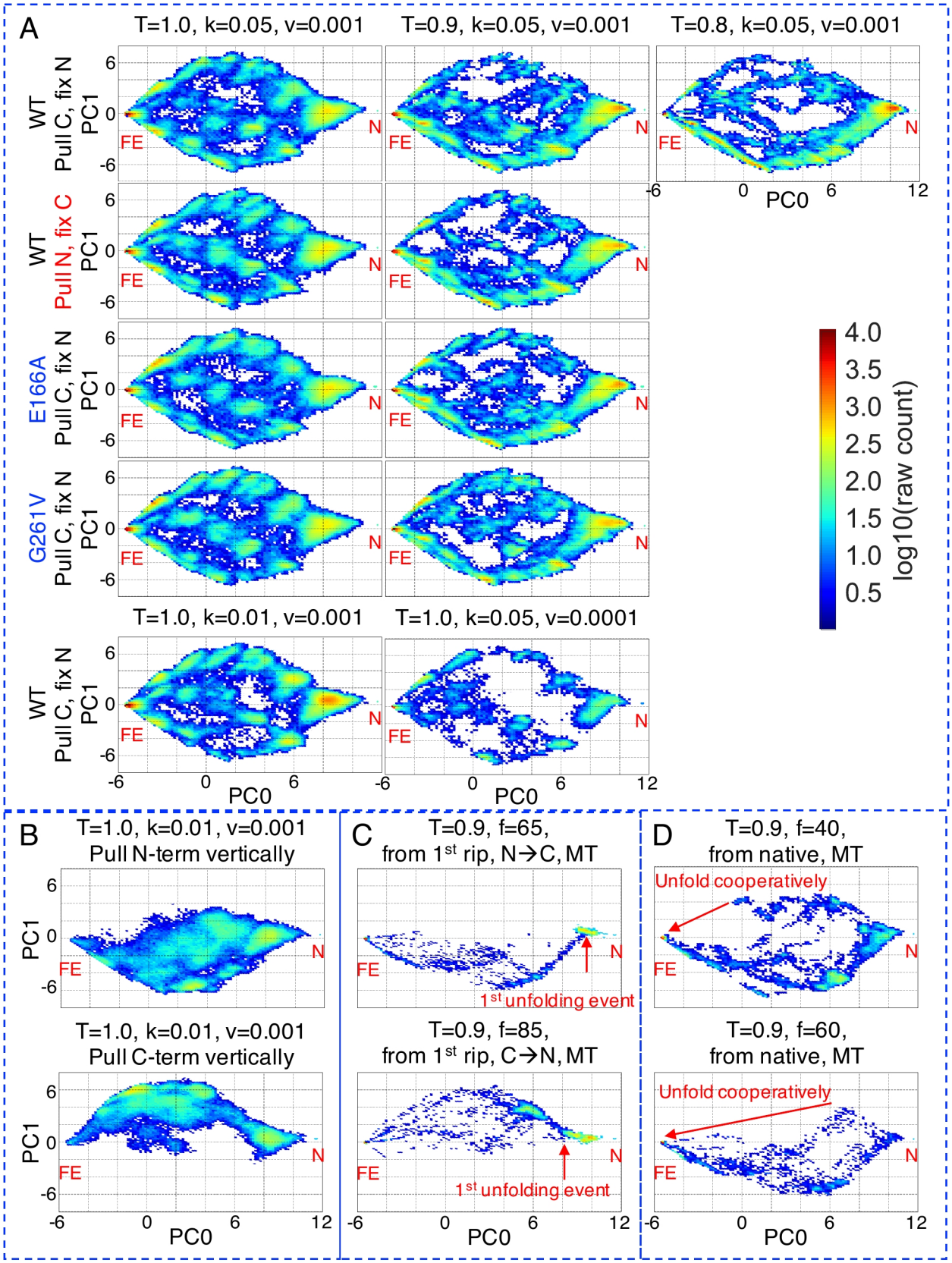
A
T=1.0, k=0.05, v=0.001
T=0.9, k=0.05, v=0.001
T=0.8, k=0.05, v=0.001
WT Pull C, fix N
WT Pull N, fix C
E166A Pull C, fix N
G261V Pull C, fix N
PC1
PC0
FE
N
log10(raw count)
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
T=1.0, k=0.01, v=0.001
T=1.0, k=0.05, v=0.0001
B
T=1.0, k=0.01, v=0.001
Pull N-term vertically
T=1.0, k=0.01, v=0.001
Pull C-term vertically
C
T=0.9, f=65,
from 1st rip, N→C, MT
1st unfolding event
T=0.9, f=85,
from 1st rip, C→N, MT
1st unfolding event
D
T=0.9, f=40,
from native, MT
Unfold cooperatively
T=0.9, f=60,
from native, MT
Unfold cooperatively

Fig. S12. Principal component analysis of unfolding trajectories of GlpG under various simulation protocols.
A. Stiff cantilever mode, pulling laterally. Each of the PCA plots is comprised of 50 trajectories, except for the trajectories using a 10x slower pulling velocity (v = 0.0001 Å/*Upside* time step), which contains only 10 trajectories. Despite differences, the fundamental heterogeneous pathway behavior remains.

B. Stiff cantilever mode, force is applied to either the N- or C-terminus vertically. The PCA plots for pulling at C- and N-terminus contains 19 or 20 trajectories. Notably, the PCA heat maps obtained in this mode fill in the blanks in the middle of the heat maps obtained in **panel A**. Those may represent structures that largely maintain the tertiary structure for the region embedded in the membrane, which would be difficult to observe in mode **A** because the tertiary structure is disrupted. Besides, the pulling the N-terminus produces "deterministically" N→C pathways as expected, and vice versa.

C. Modified MT mode simulations, pulling laterally, were re-started at the 1st unfolding event in the N→C or C→N pathway (T = 0.9 ≈ 270 K, κ = 0.05 k_BT/Å^2, v = 0.001 Å/*Upside* time step).

D. MT mode, pulling laterally, simulations were started from the native structure. The unfolding is more cooperative under higher force and in the C→N pathway than the reverse. 20 trajectories are included in each PCA plot in **panels C and D**. N and FE stand for native and fully-extended in each subplot, respectively.

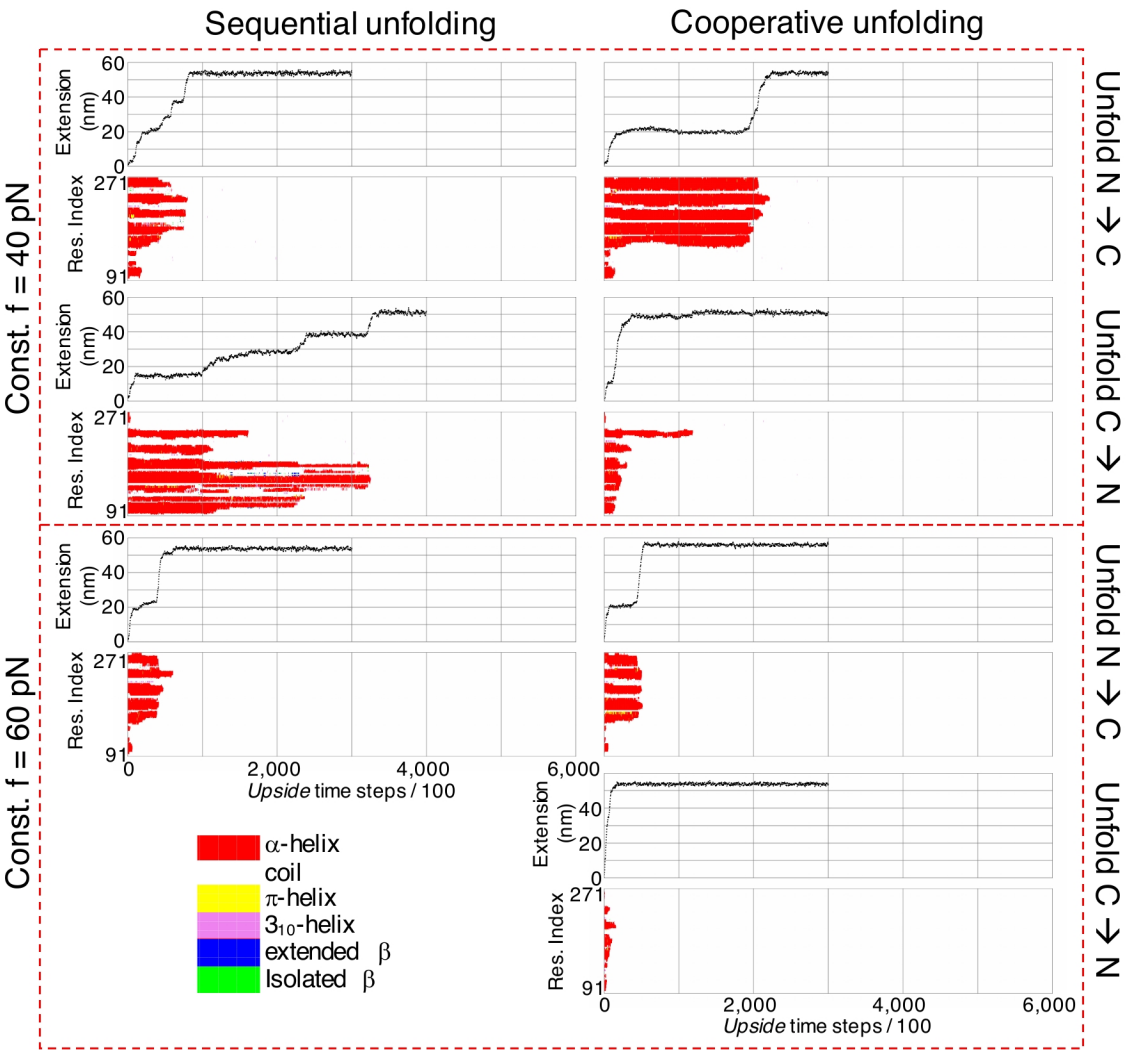

Fig. S13. Sequential and cooperative unfolding pathways of GlpG in force clamp simulations started from the native structures. Four trajectories under a constant force of 40 pN (upper panel) and three trajectories under a constant force of 60 pN (lower panel) are shown. Every trajectory is presented by a FEC and a TSS plot. Trajectories in the left column are examples of sequential unfolding pathways (with at least 3 intermediates that can be identified on the extension plot) whereas those in the right column are examples of cooperative unfolding pathways (with no more than 2 intermediates identified from the extension plot). In the 20 simulations under 60 pN, we did not observe any trajectories unfolding from C- to N-domain sequentially.

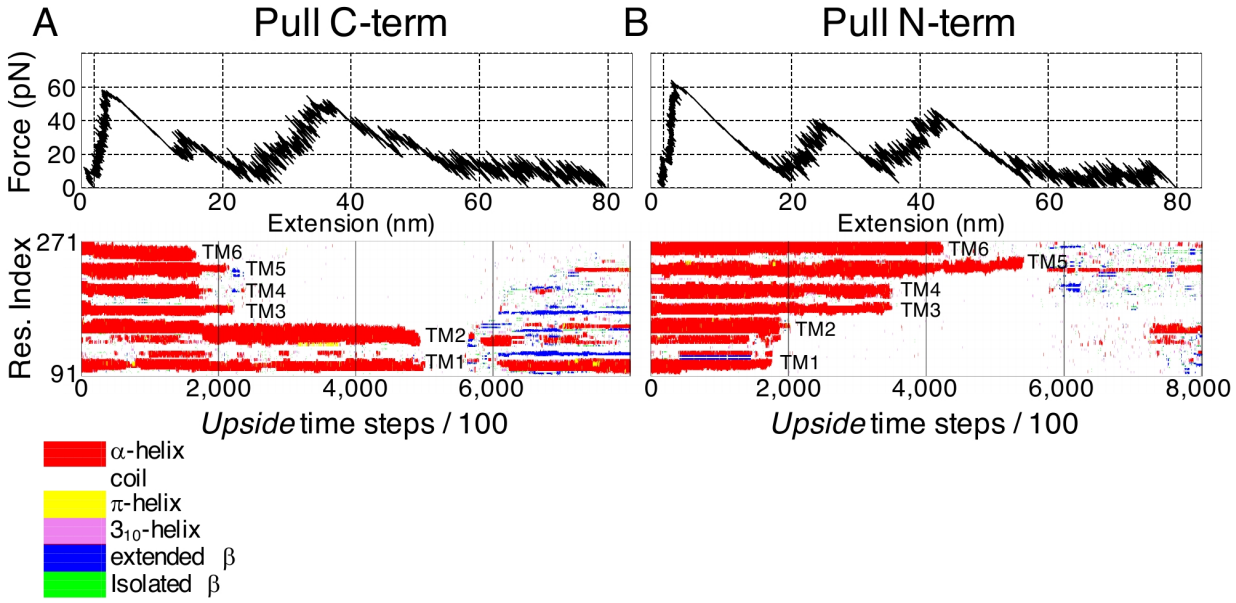

Fig. S14. Unfolding GlpG by pulling vertically in AFM, stiff cantilever mode. A, B. FEC and TSS plots of an example trajectory pulling from the C- and N-terminus, respectively. 20 simulations were performed in each case. We observed that all TM helices become completely unfolded in 4 and 3 trajectories when pulling on the C- and N-terminus, respectively. After all the protein is pulled out, the extended chain starts collapsing and forms H-bonds again. Notably, TM6 unfolds before TM5 (**panel B**), implying that TM6 is not very stable by itself in the lipid bilayer.

Table S1. Comparison of the mean unfolding force (in pN) and the s.e.m. for bR intermediates observed in experiment (13) and simulations. The spring constant (κ) is in $k_BT/Å^2$; the pulling velocity (v) is in Å/*Upside* time step; and the temperature (T) is in *Upside* temperature unit (1 ≈ 300 K). Numbers in parenthesis in the 1st column indicate the number of trajectories in which we observe all 7 helices are unfolded in the order ED, CB, and A versus the total number of trajectories we have simulated. The major intermediate identified in each region is put in the parentheses, as indexed by the last folded residue, in the 2nd to 4th columns. The two rows highlighted in light green are the results presented in the main text.

	Major intermediate in ED helix pair (residue index)	Major intermediate in CB helix pair (residue index)	Major intermediate in helix A (residue index)
Experiment	94±1 (A160)	49±2 (V101)	62±0.6 (V29)
κ = 0.05, v = 0.1 (16 / 20)	218.5±3.8 (F153)	311.4±4.2 (L99)	443.8±5.8 (V29)
κ = 0.05, v = 0.01 (11 / 20)	83.9±6.4 (F153)	70.2±5.8 (L99)	61.5±3.5 (V29)
κ = 0.05, v = 0.001 (48 / 90)	82.7±2.4 (F153 [a])	43.6±2.1 (L100)	22.6±1.6 (V29)
κ = 0.05, v = 0.0005 (25 / 60)	75.9±1.1 (F153)	40.0±1.9 (L100)	18.9±2.1 (V29)
κ = 0.05, v = 0.0001 (10 / 20)	64.5±4.0 (F153)	33.8±2.1 (L100)	7.2±1.4 (V29)
κ = 0.01, v = 0.001 (46 / 80)	69.7±1.4 (F153)	41.2±1.0 (L100)	22.5±0.9 (V29)
κ = 0.05, v = 0.001, trimer (43 / 45)	88.7±3.0 (F153 [a])	74.5±4.4 (L100)	57.5±2.7 (V29)
κ = 0.05, v = 0.001, MPx2 [b] (20 / 20)	91.5±3.3 (F153)	51.4±3.0 (L100)	38.9±2.2 (V29)

a. The comparison between experiment and simulations is conducted for the most populated intermediate, which is given in the parentheses; however, we also observe a K159 intermediate in the simulations which corresponds to the major experimental intermediate.

b. The membrane potential in this set of simulations is doubled, while all the other potentials keep unchanged. The increase of membrane potential significantly stabilizes the last helix, helix A, in the membrane bilayer, as the unfolding of helix A is observed in all 20 of 20 trajectories.

Table S2. Comparison of bR intermediates identified in the 2017 experiment (13), our monomer and trimer simulations (spring constant κ = 0.05 k_BT/Å^2, pulling velocity = 0.001 Å/*Upside* time step, T ~ 300 K) and **a 2016 CG study (21)** (see Fig. 5A in ref. (21), the intermediates are taken from the analysis of force peak groups, which were compared to previous experiments(24-26))**.** The position of an intermediate is indexed by the last folded residue of that intermediate in the protein.

	2017 experiment	Monomer simulations	Trimer simulations	2016 CG study	Description
Intermediates in helix pair ED	160	159	159		Top of helix E
	157	157	157	156.8	
	154	155, 153	155, 153		
	151	151	151	150.8	
	148	149	148		
	146	145	145		
	143	143	143		
	139	139	139	140.7	
	136		137		
	132				
	130				Bottom of helix E
	129	129			
	127	127	127		Top of helix D
	124	124	124		
	119	118	119		
		115	114		
		111	111		
Intermediates in helix pair CB	101	102, 100	100	101.4	Top of helix C
	96	97	98	95.4	
		94	94		
	91	92	92		
		88	88, 86	89.0	
	83	83	81		Bottom of helix C
	77	77		75.7	
	71				
	63	62	62		Top of helix B
		57	59, 57		
	54	54	55		
Intermediates in helix A				33.8	
	29	29	29	29.6	Top of helix A
		25	25		
			23, 21	21.2	
			18	19.1	
	16	15	15		
	8				Bottom of helix A

Table S3. Summary of unfolding pathways of GlpG.

Pulling scheme	T	κ or F [a]	WT/mutant	N- → C-domain [b]	C- → N-domain [c]
Gradual-pulling simulations from the native structure					
Pull C-term, fix N-term	1	κ = 0.05	WT	27	23
Pull C-term, fix N-term	1	κ = 0.05	E166A	27	23
Pull C-term, fix N-term	1	κ = 0.05	G261V	22	28
Pull C-term, fix N-term	1	κ = 0.01	WT	21	29
Pull C-term, fix N-term [d]	**0.9**	**κ = 0.05**	**WT**	**40**	**10**
Pull C-term, fix N-term [e]	0.9	κ = 0.05	WT	7	3
Pull C-term, fix N-term	0.9	κ = 0.05	E166A	41	9
Pull C-term, fix N-term	0.9	κ = 0.05	G261V	35	15
Pull C-term, fix N-term	0.8	κ = 0.05	WT	42	8
Pull N-term, fix C-term	1.0	κ = 0.05	WT	32	18
Pull N-term, fix C-term	0.9	κ = 0.05	WT	43	7
Force clamp simulations from the native structure [f]					
Pull N-term, fix C-term	0.9	F = 40	WT	16	4
Pull N-term, fix C-term	0.9	F = 60	WT	13	7

a. κ or F is listed as relevant to the mode of applying force, gradual pulling or force, respectively.

b. The number of trajectories with an unfolding pathway that is initiated from the N-domain. For example, in **Fig. S9**, **A**, **B** and **C** are all counted as N- to C-domain unfolding pathways.

c. The number of trajectories with an unfolding pathway that is initiated from the C-domain. For example, in **Fig. S9**, **D** and **E** are both counted as C- to N-domain unfolding pathways.

d. The primary data set that is shown in the main text.

e. The only data set of gradual pulling that uses a pulling velocity = 0.0001 Å/*Upside* time step, 10x slower than all the other data sets of gradual pulling.

f. "From the native structure" refers to starting the simulation from the native structure instead of an intermediate at the first unfolding event.

Table S4. Contour length (Lc) of bR intermediates. The truncated bR is named after the structural position (i.e., residue index) of the intermediate, as defined by the last folded residue. Numbers in parentheses indicate the number of residues of the truncated bR molecules. For example, A160 has 72 residues. The Lc of the truncated bR in its fully extended state in simulation, and the Lc of unfolded segment of the corresponding intermediate in experiment are listed.

Truncated bR	Lc, simulation (nm) (Fitted with Lp = 0.4 nm)	Lc, experiment (nm)	Description
A160 (72)	28.4	26.9	Top of helix E
T157 (75)	29.2	27.6	
F154 (78)	30.1	28.1	
V151 (81)	30.9	28.8	
I148 (84)	31.5	29.3	
L146 (86)	32.0	29.9	
A143 (89)	32.7	30.5	
A139 (93)	33.7	31.3	
V136 (96)	34.4	32.0	
S132 (100)	35.2	32.7	
V130 (102)	35.7	33.3	Bottom of helix E
K129 (103)	36.1	34.0	
L127 (105)	37.5	34.9	Top of helix D
V124 (108)	39.2	36.3	
I119 (113)	42.1	38.8	
V101 (131)	52.0	48.2	Top of helix C
D96 (136)	53.2	49.6	
P91 (141)	54.3	50.6	
Y83 (149)	56.1	52.1	Bottom of helix C
P77 (155)	57.0	53.9	
F71 (161)	59.0	55.8	
G63 (169)	62.8	58.8	Top of helix B
F54 (178)	68.3	63.3	
V29 (203)	80.1	74.6	Top of helix A
G16 (216)	83.2	77.2	
P8 (224)	84.5	79.3	Bottom of helix A

Table S5. Simulation details. Parameters used in the *Upside* simulation are summarized in the table. Units for κ, v, F, and T are the same as in **Table S2**.

System	Cantilever[a]	Attachments	κ	v	F	T	Number of simulations
Ubiquitin (fully-extended) [b]	Soft	Pull both termini in opposite direction			0, 10, 30, 50, 100, 250, 800	1.0	1 per F
Truncated bR species	Stiff	Pull C-term vertically [c], fix N-term [d]	0.05	0.001		1.0	1 per species
bR	Stiff	Pull C-term vertically	0.05	0.1		1.0	20
bR	Stiff	Pull C-term vertically	0.05	0.01		1.0	20
bR	Stiff	Pull C-term vertically	0.05	0.001		1.0	90
bR MPx2 [e]	Stiff	Pull C-term vertically	0.05	0.001		1.0	20
bR	Stiff	Pull C-term vertically	0.01	0.001		1.0	80
bR	Stiff	Pull C-term vertically	0.05	0.0005		1.0	60
bR	Stiff	Pull C-term vertically	0.05	0.0001		1.0	20
bR trimer [f]	Stiff	Pull C-term vertically	0.05	0.001		1.0	45

(To be continued in the next page.)

System	Cantilever[a]	Attachments	κ	v	F	T	Number of simulations
GlpG	Stiff	Pull C-term laterally [c], fix N-term	0.05	0.001		0.8, 0.9, 1.0	50 per T
GlpG	Stiff	Pull N-term laterally, fix C-term	0.05	0.001		0.9, 1.0	50 per T
GlpG	Stiff	Pull C-term laterally, fix N-term	0.05	0.0001		0.9	10
GlpG E166A	Stiff	Pull C-term laterally, fix N-term	0.05	0.001		0.9, 1.0	50 per T
GlpG G261V	Stiff	Pull C-term laterally, fix N-term	0.05	0.001		0.9, 1.0	50 per T
GlpG	Stiff	Pull C-term laterally, fix N-term	0.01	0.001		1.0	50
GlpG intermediates before 1st rip	Soft	Pull C-term laterally, fix N-term			64.6, 84.5	0.9	20 per F [g]
GlpG	Soft	Pull C-term laterally, fix N-term			40, 60	0.9	20 per F
GlpG	Stiff	Pull C-term vertically	0.01	0.001		1.0	20
GlpG	Stiff	Pull N-term vertically	0.01	0.001		1.0	20

a. Soft mode refers to the use of a very soft cantilever to mimic a magnetic tweezers measurement where the force is held essentially constant force after the first unfolding event occurs as the magnetic field varies slowly, on micron length scale, which is longer than the unfolded segments.
b. The simulations were started from a fully extended state.
c. The direction is relative to the membrane bilayer.
d. Held with an equally stiff spring.
e. The membrane potential in this set of simulations is doubled, while all the other potentials keep unchanged.
f. Chain A of the trimer is pulled, while the other two subunits of the trimer are allowed to undergo conformational changes freely.
g. One of the output file is corrupted, so there are only 19 trajectories useful for analysis.

Table S6. Inferred Lc values (in nm) associated with each residue in helices E to A of bR. Lc values obtained directly from the simulations of truncated bR (**Fig. S2**) are in red. The Lc value is for the unfolded segment C-terminal to a residue. For example, the contour length for the unfolded segment from the C-terminus to K159 (having residues 160 to 232) is 28.7 nm. Note, the last 4 residues (I4, Q3, A2, E1) are not in 1qhj.pdb, and therefore are not included in the simulations. In this case we don't observe the last intermediate A2 in the experiment (13).

Helix E	**A160**	K159	S158	**T157**	F156	G155	**F154**	F153	L152	**V151**
	28.4	28.7	28.9	**29.2**	29.5	29.8	**30.1**	30.4	30.6	**30.9**
	Y150	L149	**I148**	Y147	**L146**	M145	A144	**A143**	T142	S141
	31.1	31.3	**31.5**	31.8	**32.0**	32.2	32.5	**32.7**	33.0	33.2
	I140	**A139**	W138	W137	**V136**	F135	R134	Y133	**S132**	Y131
	33.4	**33.7**	33.9	34.2	**34.4**	34.6	34.8	35.0	**35.2**	35.4
	V130	K129	T128							
	35.7	36.1	36.8							
Helix D	**L127**	A126	G125	**V124**	L123	G122	T121	G120	**I119**	M118
	37.5	38.1	38.6	**39.2**	39.8	40.4	40.9	41.5	**42.1**	42.6
	I117	G116	D115	A114	G113	V112	L111	A110	L109	I108
	43.2	43.8	44.3	44.8	45.4	46.0	46.5	47.0	47.6	48.2
	T107	G106	Q105	D104	A103	D102				
	48.7	49.2	49.8	50.4	50.9	51.4				
Helix C	**V101**	L100	L99	A98	L97	**D96**	L95	L94	L93	L92
	52.0	52.2	52.5	52.7	53.0	**53.2**	53.4	53.6	53.9	54.1
	P91	T90	T89	F88	L87	W86	D85	A84	**Y83**	R82
	54.3	54.5	54.8	55.0	55.2	55.4	55.6	55.9	**56.1**	56.2
	A81	W80								
	56.4	56.6								
Linker btw C and B	Y79	I78	**P77**	N76	Q75	E74	G73	G72	**F71**	P70
	56.7	56.8	**57.0**	57.3	57.7	58.0	58.3	58.7	**59.0**	59.5
	V69	M68	T67	L66	G65	Y64				
	60.0	60.4	60.9	61.4	61.8	62.3				
Helix B	**G63**	L62	L61	M60	S59	L58	Y57	M56	T55	**F54**
	62.8	63.4	64.0	64.6	65.2	65.9	66.5	67.1	67.7	**68.3**
	A53	I52	A51	P50	V49	L48	T47	T46	I45	A44
	68.8	69.2	69.7	70.2	70.7	71.1	71.6	72.1	72.5	73.0
	Y43	F42	K41	K40	A39	D38	P37			
	73.5	74.0	74.4	74.9	75.4	75.9	76.3			
Linker	D36	S35	V34	G33	M32	G31	K30			
	76.8	77.3	77.7	78.2	78.7	79.2	79.6			
Helix A	**V29**	L28	F27	Y26	L25	T24	G23	L22	G21	M20
	80.1	80.3	80.6	80.8	81.1	81.3	81.5	81.8	82.0	82.2
	L19	A18	T17	**G16**	L15	A14	L13	W12	I11	W10
	82.5	82.7	83.0	**83.2**	83.4	83.5	83.7	83.8	84.0	84.2
	E9	**P8**	R7	G6	T5	I4	Q3	**A2**	E1	
	84.3	**84.5**	84.9	85.3	85.7	Not in 1qhj.pdb				

Table S7. Parameters describing the potential barriers that stabilize secondary structural elements of bR against force. The values in comparison are the width of potential ($\mathbf{\Delta x^{\neq}}$ [Å]) from a 2004 experimental study (7). The uncertainty of the fitted width of potential is computed from the s.e.m. of the unfolding forces. In specific, mean force ± s.e.m of forces is used in equation $\boldsymbol{F_{mp} = (\beta \cdot \Delta x^{\neq})^{-1} ln \frac{\beta \cdot \Delta x^{\neq} \cdot \kappa \cdot v}{k_0}}$ where β = $(k_BT)^{-1}$ to obtain the upper and lower boundaries, $\mathbf{\Delta x^{\neq}_{upper}}$ and $\mathbf{\Delta x^{\neq}_{lower}}$, of the width of potential. The uncertainty is $\mathbf{\left[\left(\Delta x^{\neq}_{upper} - \Delta x^{\neq}\right) + \left(\Delta x^{\neq} - \Delta x^{\neq}_{lower}\right)\right]/2}$.

	Monomer simulations	2004 experiment
Helix pair ED	5.3±0.5 Å	3.2 Å
Helix pair CB	9.9±0.1 Å	8.6 Å
Helix A	6.1±0.1 Å	6.8 Å

Movie S1. Unfolding trajectories of bR and GlpG. Simulations are run with a spring constant κ = 0.05 $k_BT/Å^2$ and pulling velocity = 0.001 Å/*Upside* time step at T ≈ 300 K.
A. Force-induced unfolding of bR in the AFM (stiff spring) mode. Force is applied vertically to the membrane surface.
B. Force-induced unfolding of GlpG in the AFM (stiff spring) mode, unfolding from N→C, from C→N, and from the middle, respectively. Force is applied laterally to the membrane surface.

REFERENCES

1. Baker, R. P., and S. Urban. 2012. Architectural and thermodynamic principles underlying intramembrane protease function. Nat Chem Biol 8:759-768.
2. Levy, R., and M. Maaloum. 2002. Measuring the spring constant of atomic force microscope cantilevers: thermal fluctuations and other methods. Nanotechnology 13:33-37.
3. Bustamante, C., J. F. Marko, E. D. Siggia, and S. Smith. 1994. Entropic elasticity of lambda-phage DNA. Science 265:1599-1600.
4. Strick, T., J. F. Allemand, V. Croquette, and D. Bensimon. 2000. Twisting and stretching single DNA molecules. Prog Biophys Mol Bio 74:115-140.
5. Min, D., R. E. Jefferson, J. U. Bowie, and T. Y. Yoon. 2015. Mapping the energy landscape for second-stage folding of a single membrane protein. Nat Chem Biol 11:981-987.
6. Evans, R. A. 1966. Abramowitz M - Handbook of Mathematical Functions with Formulas Graphs and Mathematical Tables Nbs Applied Mathematics Series 55. Ieee Spectrum 3:161-&.
7. Janovjak, H., J. Struckmeier, M. Hubain, A. Kedrov, M. Kessler, and D. J. Muller. 2004. Probing the energy landscape of the membrane protein bacteriorhodopsin. Structure 12:871-879.
8. Woodside, M. T., and S. M. Block. 2014. Reconstructing folding energy landscapes by single-molecule force spectroscopy. Annu Rev Biophys 43:19-39.
9. McGibbon, R. T., K. A. Beauchamp, M. P. Harrigan, C. Klein, J. M. Swails, C. X. Hernandez, C. R. Schwantes, L. P. Wang, T. J. Lane, and V. S. Pande. 2015. MDTraj: A Modern Open Library for the Analysis of Molecular Dynamics Trajectories. Biophys J 109:1528-1532.

10. Kabsch, W., and C. Sander. 1983. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. Biopolymers 22:2577-2637.
11. Savitzky, A., and M. J. E. Golay. 1964. Smoothing + Differentiation of Data by Simplified Least Squares Procedures. Anal Chem 36:1627-&.
12. Oliphant, T. E. 2007. Python for scientific computing. Comput Sci Eng 9:10-20.
13. Yu, H., M. G. Siewny, D. T. Edwards, A. W. Sanders, and T. T. Perkins. 2017. Hidden dynamics in the unfolding of individual bacteriorhodopsin proteins. Science 355:945-950.
14. Muller, D. J., M. Kessler, F. Oesterhelt, C. Moller, D. Oesterhelt, and H. Gaub. 2002. Stability of bacteriorhodopsin alpha-helices and loops analyzed by single-molecule force spectroscopy. Biophys J 83:3578-3588.
15. Ainavarapu, S. R., J. Brujic, H. H. Huang, A. P. Wiita, H. Lu, L. Li, K. A. Walther, M. Carrion-Vazquez, H. Li, and J. M. Fernandez. 2007. Contour length and refolding rate of a small protein controlled by engineered disulfide bonds. Biophys J 92:225-233.
16. Stirnemann, G., D. Giganti, J. M. Fernandez, and B. J. Berne. 2013. Elasticity, structure, and relaxation of extended proteins under force. Proc Natl Acad Sci U S A 110:3847-3852.
17. Pedregosa, F., G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. 2011. Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research 12:2825--2830.
18. Jumper, J. M., N. F. Faruk, K. F. Freed, and T. R. Sosnick. 2018. Trajectory-based training enables protein simulations with accurate folding and Boltzmann ensembles in cpu-hours. PLoS Comput Biol 14:e1006578.
19. Jumper, J. M., N. F. Faruk, K. F. Freed, and T. R. Sosnick. 2018. Accurate calculation of side chain packing and free energy with applications to protein molecular dynamics. PLoS Comput Biol 14:e1006342.
20. Wang, Z., J. M. Jumper, S. Wang, K. F. Freed, and T. R. Sosnick. 2018. A Membrane Burial Potential with H-Bonds and Applications to Curved Membranes and Fast Simulations. Biophys J.
21. Yamada, T., T. Yamato, and S. Mitaku. 2016. Forced Unfolding Mechanism of Bacteriorhodopsin as Revealed by Coarse-Grained Molecular Dynamics. Biophys J 111:2086-2098.
22. Subramaniam, S., and R. Henderson. 2000. Molecular mechanism of vectorial proton translocation by bacteriorhodopsin. Nature 406:653-657.
23. Mackerell, A. D., Jr., M. Feig, and C. L. Brooks, 3rd. 2004. Extending the treatment of backbone energetics in protein force fields: limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. J Comput Chem 25:1400-1415.
24. Kessler, M., and H. E. Gaub. 2006. Unfolding barriers in bacteriorhodopsin probed from the cytoplasmic and the extracellular side by AFM. Structure 14:521-527.
25. Sapra, K. T., J. Doehner, V. Renugopalakrishnan, E. Padros, and D. J. Muller. 2008. Role of extracellular glutamic acids in the stability and energy landscape of bacteriorhodopsin. Biophys J 95:3407-3418.
26. Voitchovsky, K., S. A. Contera, and J. F. Ryan. 2007. Electrostatic and steric interactions determine bacteriorhodopsin single-molecule biomechanics. Biophys J 93:2024-2037.